Title
Effects of peer comparisons on low-promotability tasks: Evidence from a university field experiment

Authors
Villas-Boas, SB
Taylor, RLC
Deakin, E

Publication Date
2019-02-01

DOI
10.1016/j.jebo.2018.12.005

Peer reviewed

Effects of Peer Comparisons on Low-Promotability Tasks: Evidence from a University Field Experiment

Elizabeth Deakin[a], Rebecca L.C. Taylor[b], and Sofia B. Villas-Boas*[c]

[a]*Department of City & Regional Planning; University of California, Berkeley*

[b]*School of Economics; University of Sydney*

[c]*Department of Agricultural & Resource Economics; University of California, Berkeley*

November 27, 2018

Abstract

Governance—the way rules are set and implemented—in many institutions is sustained through the service of groups of individuals, performing low-promotability tasks. For instance, the success of not-for-profit professional societies, civic organizations, and public universities depends on the willingness of members and employees to serve in governance. Typically service is requested by annual calls to serve. We implement and analyze a field experiment at a large public university using a randomized experimental design, to investigate whether responses to calls to serve are affected by revealing a department's service rankings among its peer departments. We find that revealing a service ranking in the lowest quartile leads to significantly higher response rates than disclosing a median and higher-than-median ranking. Second, beyond informing department heads of their departments' service rank, directly informing individual faculty members does not have an additional effect on response rates. Third, we show that the treatment effects in the lowest serving quartile are driven by female faculty responses, even though female faculty members were no more likely than their male peers to respond to serve before the experiment. If taking on such tasks is detrimental to promotion, while important for the overall institution, this has implications for the faculty careers of women and men. Given our data we cannot identify potential mechanisms behind the results; formally testing these mechanisms is an area for future research.

Keywords: Field experiment, Social comparison, Calls to service, Low-promotability tasks, Gender differences.

JEL Classification: D64; C93; H41; D71; D03

* *Corresponding author*: Villas-Boas. *Postal address*: Department of Agricultural & Resource Economics, University of California at Berkeley, 207 Giannini Hall, Berkeley, CA 94720-3310, USA. *E-mail address*: sberto@berkeley.edu. We are grateful to the editor Daniel Houser and two anonymous referees for invaluable suggestions. We thank Peter Berck, Stefano DellaVigna, Ilyana Kuziemko, Andrew Stevens, Molly Van Dop, and seminar participants at U.C. Berkeley and at the Behavioural Economics Foundations and Applied Research Workshop at the University of Sydney, for helpful comments. We thank the university in this study for sharing data and allowing us to conduct a field experiment. All errors are our own.

1 Introduction

The principle of shared governance, while contributing to the excellence of many public universities, depends on the alacrity of faculty members to exercise it. By serving on academic senate committees, individual faculty members dedicate time to participate in the way rules are decided and implemented, and consequently benefit from having a voice in discussions affecting their lives as teachers, researchers, and employees. At the department level, a department can have its interests represented by having a large share of its faculty serving on committees. However, faculty may have reasons not to serve. Due to the opportunity cost of time, faculty may prefer that service be completed by someone else, especially if the task is seen as having little effect on their evaluation and career advancement. For instance, research-related tasks may be viewed as more promotable than service-related tasks in research-oriented universities (Babcock et al., 2017). Thus it is perhaps unsurprising that despite the reasons to serve, calls to serve on academic senate committees have low response rates.[1]

This paper investigates whether providing information to departments about their service ranking—in terms of previous service on academic senate committees relative to peer departments of the same size—changes individual faculty members' probability to answer calls to serve on future senate committees. We also investigate whether informing faculty directly about their department's service rankings (in a bottom-up fashion) instead of only

[1] This paper explores a university setting where the baseline response rate to academic senate service calls is 8.5% while a response rate of at least 18% is required to fill all positions. At a different public university, Babcock et al. (2017) find a volunteer response rate of 3.7% to academic senate service calls.

informing department heads (in a top-down fashion) leads to different responses rates. Finally, we examine whether the peer comparison treatments differentially affect the response rates of female faculty relative to their male peers.

Our experimental design is as follows. For the 64 departments and schools in our sample, we divide departments into four size quartiles. Given data on faculty senate service in the previous year, we compute the rates of service by department as well as each department's service ranking within its size quartile. We then randomly assign half of departments in each size quartile into treatment and control groups. The departments in the control group receive the standard call-to-serve reminder email, sent to department chairs or school deans every year, requesting them to forward the email to their faculty members. Clicking on the link provided in the email leads to an online form for indicating willingness to serve on a committee in the next year.[2] The treatment group receives the same direct or chair email reminder, but additionally, the departments in this group receive information on their department's service ranking among all departments of a similar size.

We collect two years of data on faculty responses to the call-to-serve email—from before and after the experiment—in order to estimate the effect of the information treatment on individual voluntary service responses. The highest serving departments in each size group are not treated due to institutional restrictions imposed at the time of the design. This leaves us with 24 treated and 24 control departments, comparable in size and service, and 16 very high service departments that receive the control email but are not used to identify the treatment effects. As an orthogonal sub-treatment to the overall experimental design,

[2] The online form also asks you to rank which committees you would like to serve on. The form does not provide a way to indicate you are not interested in serving; thus all that reply are interested in serving.

all faculty in a subset of treated and control departments receive the same call-to-serve email directly, and at the same time as their chair. In this sub-treatment, we aim to assess whether, by informing faculty directly of their department's relative ranking, this bottom-up information approach leads to different response rates as to when the chair receives the information and forwards it to their faculty in a top-down manner. This sub-treatment is also motivated by the fact that we cannot directly verify whether chairs forward the email to all of their faculty, even though the email explicitly asks them to do so.

Changing behavior due to learning the behavior of others is consistent with four distinct hypotheses (Ayres et al., 2013). First, learning that peer departments provide less (more) service could increase (decrease) feelings of guilt and lead to more (less) service in the future. Alternatively, learning the behavior of peers might provide information about the possibility of alternative time use choices of the faculty body and the relative benefits of those choices. Third, information about the volunteer rates of others may also allow individuals to update their beliefs about the equilibria being played and about the pivotality of their individual decision to volunteer. Fourth, learning how your department ranks compared to peer department may also reveal (or make salient) that your behavior is being observed and scrutinized by the university (i.e., big brother is watching), leading to a Hawthorne effect. Like other peer information studies, an important limitation of our experiment is that we are not able to distinguish between social learning, conditional cooperation, equilibrium selection, and observability theories of behavioral change. For instance, if faculty members react to information that their department is serving more than their peers' by decreasing their own service, this change might be caused either by decreased guilt (i.e., conditional cooperation), or by making a Bayesian inference about being free-ridden by others and missing out on research time (i.e., social learning and equilibrium selection). It also could be due to faculty believing this is the response desired by the university (i.e., Hawthorne effect). Another limitation of our design is that, when the call-to-serve emails go to the chair, we can not disentangle whether the treatment effect is driven by the behavior of the chair or the behavior of the individual faculty. For instance, a chair that receives a treatment email may change how they ask their faculty to serve. While we attempt to address this concern by sending call-to-serve emails directly to faculty, we are cautious in interpreting the mechanisms behind our results for this reason.

Finally, using characteristics of individual faculty, we investigate treatment effect heterogeneity by the gender of the faculty member and by whether the faculty member replied to the call-to-serve email in the previous year. Recent research identified gender differences in the probability of taking on service tasks with low promotability (Babcock et al. 2017; Babcock, Recalde, and Vesterlund 2017), finding that women are both more likely to volunteer to serve than men and are more likely to be asked to serve than men. Thus, if the treatment effect is driven by the behavior of chairs, chairs that learn their department is low performing may be more likely to ask women, as well as those that have served in the past, to serve in the future. Croson and Gneezy (2009) review the experimental evidence on preference differences between genders and find that women are neither more nor less socially oriented, but that social preferences of women are more situationally specific and malleable than those of men. Thus, if the treatment effect is driven by the behavior of individual faculty, women may be more likely than men to shift their preferences for voluntary service when provided with new information. Unfortunately, for the same reasons as above, we will not be able to disentangle these mechanisms.

Our findings are as follows. The average probability of replying to the call-to-serve email is greater with the treatment email relative to the standard control email, but not significantly. The average probability of replying also did not change significantly when faculty were directly sent the treatment email. When we go beyond the average effects and investigate heterogeneity of responses by service quartile disclosed in the treatment email, we find interesting and statistically significant results. First, the probability of responding to the call-to-serve increased the most for individuals in departments who received an email disclosing they were in the lowest service quartile. Second, in terms of gender differences, we find the treatment effects by service quartile are larger for female faculty. This is particularly interesting since, in our sample, female faculty were no more likely than their male peers to respond to the call-to-serve before the information intervention. We also find larger treatment effects for those who responded to the call-to-serve in the previous year. Thus while social comparisons increase the service response rate of the lowest serving departments, this finding is driven by the responses of female faculty and faculty already engaged in service.

While we are not able to explain why this happens, our results highlight the power of peer comparisons to influence service rates for non-promotability tasks, especially among those revealed to be behind their peers. Existing research has empirically established a link between social comparisons and other pro-social behaviors, including home electricity use (Allcott, 2011; Ayres et al., 2013), restaurant hygiene (Jin and Leslie, 2003), pollution emissions (Blackman et al., 2004), charitable giving (Frey and Meier, 2004; Smith et al., 2015), voter turnout (Gerber and Rogers, 2009), and towel reuse by hotel guests (Goldstein et al., 2008).[3] Our paper relates most closely to studies that examine heterogeneity in

[3]For reviews of the psychology literature on social influence and pro-social behavior, see Cialdini and Goldstein

treatment effects depending on where individuals fall in the behavior distribution. Allcott (2011) shows that by providing feedback to customers on home electricity usage relative to their neighbors, utilities can reduce energy consumption at a low cost. Moreover, the author finds heterogeneous treatment effects—households revealed to be in the highest decile of pre-treatment consumption decrease usage by 6.3%, while those revealed to be in the lowest decile decrease consumption by only 0.3%. In a university setting, Card et al. (2012) find that revealing the salary of peers causes university workers with salaries below the median for their pay unit and occupation to report lower job satisfaction, while those earning above the median report no higher satisfaction. Similar to these studies, we find little to no treatment effect for faculty in departments revealed to be just above/below median with respect to previous service participation, and a large treatment effect for faculty in departments revealed to be in the lowest service quartile.

Our paper also adds to the growing literature on gender differences in the allocation of time spent on low-promotability tasks. Survey evidence from academia has shown that female faculty, relative to their male peers, spend fewer hours in research, more time advising undergraduate students, and more time serving on department and college-level committees (Misra et al., 2012; Mitchell and Hesli, 2013; Porter, 2007). While these findings suggest gender differences in the likelihood of accepting requests to do low-promotability tasks, these survey data cannot rule out the possibility that assignments differ because of differences in the frequency with which men and women are asked to serve voluntarily. This is a concern in our setting as well, where we cannot confirm if chairs forward emails to (or discuss service responsibilities with) their department members differentially by gender. Our study attempts

(2004); Penner et al. (2005).

to address the concern that all faculty might not be asked equally by sending the call-to-serve email directly to all faculty members in a subset of departments, regardless of gender and rank. Moreover, by using administrative service and response data, we also avoid the systematic under/over reporting and recall bias of survey data.

Finally, this paper relates closely with the work of Babcock et al. (2017) and Babcock, Recalde, and Vesterlund (2017). Using controlled laboratory settings, these studies show that women are both more likely to volunteer for low-promotability tasks than men and more likely to be asked to volunteer for these tasks than men. In addition to the lab experiments, Babcock et al. (2017) examine university service data similar to the data used in this study. They test whether women in another large, public U.S. university are less able to say "no" than men to calls (sent via email) to serve on a faculty senate committee. These call-to-serve emails are sent directly to all faculty—an important contrast to our setting where the call-to-serve emails are sent to department chairs (except for the sub-treatment with direct faculty emails). We replicate the Babcock et al. (2017) analysis on our pre-experiment response data and with the post-experiment response data from the direct email treatment. While Babcock et al. (2017) find that female faculty are significantly more likely than male faculty to volunteer to be on a committee and to serve on committees, we find no difference in response and service rates between men and women in the pre-experiment (chair-only) sample or in the experimental direct-email sample. Babcock et al. (2017) state, "What is not clear from these field data is why women are more likely than men to accept such requests." We echo this and add, it is also not clear why women at one university would be more likely than men to accept service requests but not at another university. Our findings suggest that institutional norms (i.e., how faculty believe service will affect their promotion), the information departments (both chairs and faculty) have about their service relative to their peers' service, and how faculty are asked to volunteer could explain some of the gender differences in performing low-promotability tasks, and consequently, could be an important component in the career trajectory of women relative to men. To draw more general conclusions about whether and why female faculty are more likely to volunteer for low-promotability service, systematic data collection and analysis are needed on call-to-serve response rates at other institutions. One limitation of our study is that we cannot pinpoint why women respond differently than men to the treatment, or in other words, we cannot identify the mechanism behind the gender difference, as it can be due to differences in pro-social behavior by women relative to men, differences in being asked to serve, and finally, differences in being able to say no when asked.

The paper proceeds as follows. Section 2 describes the data and experimental design and section 3 lays out the empirical strategy. Section 4 presents the results, including average and heterogeneous treatment effects. Section 5 concludes and discusses avenues for future research.

2 Experimental Design and Data

2.1 Data Sources

The first dataset we use consists of a faculty census for each department and school in a large, public university in the U.S., for the academic year 2013-14. The raw data for these administrative records consist of 1,719 observations, each with a faculty identifier, department affiliation(s), the academic title (assistant, associate, etc.), and a job-code. Several job-codes in a particular school were not actual faculty senate positions, so these 109 observations were dropped. There are 69 departments in the raw data, however, five of these departments mostly teach undergraduate students in interdisciplinary units and do not have a core faculty body or chair. These departments are also dropped from the analysis. As a result of these criteria, we have a final roster of 1,501 faculty belonging to 64 departments.

The second dataset we use originates from public records and consists of the *aggregate* number of faculty serving in the senate for each department and year from 2005 to 2014. The third dataset is a list of all faculty that served on committees in 2013-14, which is matched to the faculty census. The fourth and final dataset consists of the individual faculty responses to the call-to-serve email for 2013-14 (before the treatment) and for 2014-15 (after the treatment). Every year call-to-serve emails are sent to department heads, requesting them to forward the email to their faculty. In 2013-14, the response rate to the call-to-serve email was 8.5% while a response rate of at least 18% was required to fill all positions. When there are not enough volunteers from the call-to-serve email, faculty that served in the past are asked to serve again. The necessity of relying on the same volunteers over and over and not having enough new volunteers was a concern for the academic senate and a motivating factor behind permitting this study.

Table 1 presents summary statistics for the number of faculty members by department in 2013-14, as well as on the number of service eligible faculty members by department. The service eligible pool consists of faculty of associate title or higher, as assistant professors are discouraged from serving on committees and, consequently, only associate professors or higher typically serve.[4] The departments in this institution are quite heterogeneous in

[4]Of the 211 assistant professors in our sample, only 5 responded to the 2013-14 call-to-serve email and only

their faculty size.[5] We divide the departments into size quartiles based on the number of service eligible faculty. The smallest quartile (labeled *Size Q1*, henceforth) consists of 13 departments with 7.5 or less service eligible faculty. The second quartile in terms of size (*Size Q2*) has 19 departments with 7.5 to 12 faculty members that are eligible for service. The next quartile (*Size Q3*) has 15 departments with the number of eligible faculty ranging from 12 to 24.5. Finally, the largest quartile (*Size Q4*) has the remaining 17 departments and these are departments that are quite large, with up to 72.5 faculty members in the eligible service pool.

Given the above datasets, we compute the percentage of eligible faculty that served in 2013-14 for each department. Table 1 provides a breakdown of service participation summary statistics by department size quartile in 2013-14. Average service eligible participation ranges from 22% to 33% across the size quartiles and median service eligible participation ranges from 17% to 32%. *Size Q1* has the lowest median participation at 17%. This would be consistent with faculty in very small departments being already fully involved in internal departmental level service, with little time to participate in additional campus level service. Moreover, the smaller departments (*Size Q1* and *Q2*) have the largest maximum department service participation of 75%. This is not surprising given faculty size can be as low as 2.5 in *Size Q1*, and therefore if one of those few faculty is involved in service, the resulting service

4 served on academic senate committees in 2013-14.

[5] In the raw data, several faculty had multiple affiliations; that is, they were listed as belonging to multiple departments. In order to compute the total number of faculty in each department, a faculty member was counted as $\frac{1}{n}$ of a faculty member in each of their n departments of affiliation. Thus, if a faculty member was listed as having two affiliations (which happened in 73 cases), those faculty received a $\frac{1}{2}$ count in each department.

participation rate is quite high. This is why it is important to analyze and compare service participation rates among departments of similar size. And we do this by randomly revealing to departments their service rankings *within* each column, not across columns.

2.2 Experimental Design

Within each department size quartile (*Size Q1-Q4*), we divide departments into four service quartiles based on their eligible service participation rate in 2013-14: *Serve Q1* $\leq$ 25th percentile in terms of service; *Serve Q2* = 26-50th service percentile; *Serve Q3* = 51-75th service percentile; and *Serve Q4* $\geq$ 76th service percentile. Next, within each department size quartile, we randomly assign departments into treatment and control groups.[6] We do not assign treatment to any departments in *Serve Q4* due to institutional restrictions. In the original research design, *Serve Q4* departments also were assigned into treatment and control groups, to test whether disclosing high service would encourage or discourage future service responses. Both conditional cooperation and social learning mechanisms predict a "descriptive norm boomerang effect"—departments with high pre-treatment service should decrease service, while those with low pre-treatment service should increase service (Allcott, 2011). However, we did not find institutional support from the data source to implement this additional analysis. We opted to follow their recommendation, given their willingness to share these unique data made this study possible in the first place. We therefore only make an experimental intervention on the three lowest service quartiles and leave the call-to-serve email to *Serve Q4* departments unchanged. Table 2 presents the number of departments in

[6]The randomization is not stratified on service quartile.

each size quartile cross-tabulated with each service quartile and treatment group.[7]

The control department chairs receive a standard call-to-serve email from the chair of the committee that recruits service on campus. This email is similar to the call-to-serve email they receive every year and asks chairs to forward the email to their faculty. The treated department chairs receive the same email but with additional information on where their department ranks in terms of service. In particular, the treated email reveals the quartile of the department in terms of service participation among departments of similar size (i.e., in the same size quartile). Given that we do not treat the highest service quartile departments, the tone of both the treatment and control emails encourages departments to improve their participation rate in the future. Finally, we randomly select a subset of treated and control departments to which we send emails directly to the faculty in addition to the emails sent to their chair.

The standard email sent to the chairs each year starts with "Dear Chair, We would like to thank your department for its ongoing participation in the activities of the [University's] Academic Senate."[8] The email goes on to discuss several reasons why it is in the interest of the faculty members to undertake Academic Senate service. The email concludes with, "Please forward this message to your faculty to encourage them to submit their committee preferences directly. They can sign up by following this link." In the treatment email, we add an additional line saying, "After reviewing the service participation data for 64 campus

---

[7]For *Size Q1*, nearly half of the departments had zero service participation. Instead of grouping some of these zero service departments into *Serve Q1* and some into *Serve Q2*, we put all of them in *Serve Q1*, which means there are no *Size Q1* departments in *Serve Q2*.

[8]Redacted versions of the call-to-serve emails are presented in the appendix.

units, we have noted that your department is ____ when compared to other units of similar size.” The blank space varies according to the service quartile of each department, between, “in the bottom 25%”, “below the median (in the 25%-50% range)”, and “above the median (in the 50%-75% range)”. Finally, all faculty in a random subset of treated and control departments receive the same call-to-serve email directly, and at the same time as their chair. In this sub-treatment, we wish to assess whether a bottom-up information approach leads to different response rates than the university’s usual top-down approach.

We have two testable hypotheses from this experimental design. First, we expect larger treatment effects (i.e., increases in response rates) for those revealed to be in the lowest performing departments versus those revealed to be in median performing departments. This hypothesis is consistent with several mechanisms—conditional cooperation, social learning, equilibrium selection, and observability effects—however, our experimental design is unable to distinguish between these mechanisms. Second, if a stronger signal comes from direct emails versus emails forwarded from a department chair—either because the chair does not forward the email or because a forwarded email seems less important—we would expect larger treatment effects for those that get direct emails than for those that get emails forwarded from their chair. Conversely, if receiving a direct email sends a weaker signal than one from a department chair—perhaps because the recipient does not personally know the sender or because direct emails from the faculty senate got lost among the plethora of university emails received—we would expect larger treatment effects for those that get emails from the chair.

2.3 *Concerns for Interpretation*

An important concern when interpreting the results of our experiment is whether chairs actually forward the call-to-serve emails. Even though the email explicitly asks chairs to forward the email to their department, we cannot verify that chairs do so, just as we cannot verify that faculty read their emails. If we find a treatment effect, it could be that chairs are more likely to forward the email to their faculty when they receive a treatment email. In other words, the treatment effect we estimate could have more to do with how the treatment influences chairs rather than the general population of faculty members.

In order to address this concern, we implement a follow-up survey of the 64 chairs in our sample.[9] When asked how they forwarded call-to-serve emails inviting volunteers for academic senate service each year, 67% of chairs respond that they forward the emails to all faculty in their department, 13% respond that they never forward the emails, and 13% respond that they selectively forward emails depending on the service availability and likelihood to serve of their faculty. When asked how their department decides who volunteers to serve on academic senate committees each year, a vast majority of chairs (83%) respond that their faculty individually decide whether or not to serve. Importantly, the proportion of these responses does not differ by whether the chair was in a treated or control department.[10] While this survey captures stated behaviors and not actual behaviors, we find the responses reassuring in that most chairs are not deciding to how many (and which) faculty members to forward the email and "volunteer."

[9] We received responses from 30 chairs, a response rate of 48%.

[10] One third of both treated and control chairs in *Serve Q1* report selectively forwarding emails to their faculty.

2.4 Balance in Pre-Treatment Observables for Treated and Control Departments

To validate this empirical design, it is important to investigate whether the control and treatment groups are indeed comparable in attributes—with respect to the number of faculty, the number of departments, and average service in the pre-period. Table 3 presents the summary statistics of observable characteristics for the entire sample of departments (column 1) and by service quartile (columns 2–5). Additionally, Table 3 breaks down the summary statistics by treatment and sub-treatment groups (columns 6–9). To make visual comparison easier, bold text in columns 2–4 indicates the average for *Serve Q1*, *Q2*, or *Q3* is statistically different (with a p-value ≤ 0.05) from the average for *Serve Q4* (column 5). Bold text in columns 6 and 8 indicates the average for the treated departments is statistically different (with a p-value ≤ 0.05) from the average for the corresponding control departments in columns 7 and 9. In total, there are 24 treated departments with 557 faculty, 24 control departments with 572 faculty, and 16 *Serve Q4* departments with 375 faculty members. With respect to the sub-treatment, 6 of the treated department and 6 of the control departments were randomly selected to receive direct faculty emails.

Considering the whole sample, the average department has approximately 23 faculty—a third of which are female—a response rate to the pre-treatment call-to-serve email of 9%, and an Academic Senate service rate of 18%. While the female share of faculty increases with service quartile from 31% in *Serve Q1* to 39% in *Serve Q4*—suggesting higher serving departments have a higher share of women, these averages are not statistically different from each other. As expected by design, call-to-serve response rates and service rates in 2013-14 increase with service quartile. The response rates are 3%, 8%, 12%, and 13% for *Serve Q1*, *Q2*, *Q3*, and *Q4* respectively, and the corresponding service rates are 7%, 19%, 20%, and 27%.

In terms of pre-existing trends in service participation since 2005, service in *Serve Q1* departments had been declining by 2 percentage point per year while service rates in the other three quartiles had been constant or increasing over the same time.[11] Turning to the treatment groups, it is reassuring to find almost no statistically significant differences between treated and control departments.[12] The notable exception is that the control departments in the direct email sub-treatment have a higher service rate in 2013-14 than the corresponding treated departments.

To investigate pre-existing trends in the percentage of women in the department faculty, we use data on the number of women and the total number of faculty by department from 2008 to 2014, to construct the percentage women by department by year. We see in the last row that the percentage of women had been overall constant, given the non-significant estimated linear trend coefficient, as well as constant for all quartiles and when broken down by control and treated departments.

Of particular note, when we examine the service rates across gender, we find no statistically significant difference between women and men in 2013-14. This is true across all columns (i.e., for the whole sample and all quartile and treatment groups). Conversely, when we compare service rates within gender, women in *Serve Q4* serve at a higher rate than women in *Serve Q1*, *Q2*, and *Q3*, though not significantly so for *Serve Q2*, whereas

[11] The average trend in service participation is calculated using the dataset, from public records, consisting of the aggregate number of faculty serving in the academic senate for each department and year from 2005–2014.

[12] This is particularly reassuring given we did not stratify the randomization on service quartile.

men in *Serve Q4* serve at a significantly higher rate than men in *Serve Q1*, but not men in *Serve Q2* and *Q3*. These summary statistics provide weak evidence that being a high service department is correlated with having a higher share of women and a higher share of women serving. In other words, the service rankings revealed to departments may be more correlated with the service rates of women than men.

In summary, Table 3 shows that treated and control departments are well balanced with respect to observable characteristics and pre-trends in service, and percentage of women in their faculty, lending validity to our randomization and experimental design. Furthermore, on average we cannot conclude that women serve at a higher rate than men in our baseline year.

2.5 Post-Treatment Call-to-Serve Response Rates

A total of 233 responses were obtained for the 2014-15 call-to-serve emails. Compared to the previous year's 127 responses, the response rate increased by 83%.[13] The top panel of Table 4 breaks up responses and non-responses by treatment group. We include all ranks of faculty in this table, given that we care about willingness to respond to the call-to-serve, not whether a faculty member can actually serve.[14] The total number of faculty that did and did not respond are reported separately for both treated and control departments whose chair received the call-to-serve email (titled "Chair Only Email") and departments whose faculty directly received the call-to-serve email (titled "Direct Faculty Email"). The control rows

[13] While the aggregate data on department service goes back to 2005, we only have 2013-14 and 2014-15 response data at the individual faculty level.

[14] The institution discourages assistant professors from serving until tenure, which means only associate professors or higher can serve.

exclude observations originating from the *Serve Q4* departments. We find that the treated Chair Only Email group had a response rate of 16% while the corresponding control group had a response rate of 13%. In terms of the Direct Faculty Email, the response rate for the treated group is 23% and the response rate for the control group is 19%. Thus a preliminary look at the response data suggests receiving a treatment email leads to higher response rates.

The bottom panels of Table 4 report heterogeneity in response rates for treated and control faculty by their departments' 2013-14 service quartile, by their gender, and by whether faculty members replied to the call-to-serve email in the previous year. For *Serve Q1*, the treated group response rate is twice as large as the response rate of the control group (19% versus 9%). Conversely, for *Serve Q2* and *Serve Q3*, the response rates for the treated groups are smaller than the corresponding control groups. Both treated men and women have higher response rates than their control counterparts—the share of men responding is 4 percentage points higher in treated departments than control departments and the share of women responding is 1 percentage point higher. Control faculty members that responded in 2013-14 have a much higher response rate in 2014-15 than those that did not previously respond (51% vs 11%) and these rates are slightly greater for the corresponding treated faculty (53% vs 14%).

A visual depiction of the call-to-serve replies by department given replies in the previous year is presented in Figure 1. On the horizontal axis, we have the share of faculty by department responding to the call in 2013-14, and on the vertical axis, the share of faculty by department responding to the call in 2014-15. Each circle is a department with the size of the circle corresponding to the size of the department. In panel (a), we have the control departments depicted in blue, and in panel (b), the treated departments are depicted in red. The circles above the 45-degree line are departments where there was an increase in responses relative to the previous year. In panel (a), we see most control departments are near the 45-degree line, indicating response rates did not change much from the previous year. Conversely, in panel (b) the treated departments are dispersed farther above the 45-degree line than the control departments, suggesting the treatment email had a positive average effect on responding to the call-to-serve.

Finally, Figure 2 compares the response and service rates pre- and post-treatment, for all departments and by service quartile. Panel (a) reveals the share of faculty responding to the call-to-serve email increased post-treatment for all departments, and especially for the departments that were the lowest serving pre-treatment. Moreover, the response rates for 2014-15 are nearly equal across service quartiles, averaging 16%. This is much closer to the 18% response rate needed to fill all committee positions in the academic senate than the average in 2013-14. Panel (b) shows that while the average service rate across all departments does not change post-treatment, the share of faculty serving from *Serve Q1* increases while the share serving in *Serve Q4* decreases post-treatment. This is consistent with the treatment redistributing the service load from the high to low serving departments through who responds to the call-to-serve. In summary, the above tables and figures provide initial evidence that the social comparison treatment altered the distribution of departments responding to the call-to-serve email, which we investigate formally in the following regression analysis.

3 Empirical Model

The probability of responding to the 2014-15 call-to-serve among faculty in our 2×2 treatment design is modeled as:

$$(1) \qquad R_{fd} = \alpha_0 + \gamma_1 Treat_d + \gamma_2 DirectEmail_d + \gamma_3 Treat_d \times DirectEmail_d + X\beta + \epsilon_{fd}$$

where R_{fd} is a response indicator equal to 1 if faculty member f from department d responds to the 2014-15 call-to-serve, and equal to 0 otherwise. Coefficient α_0 captures the average response rate without either treatment. The coefficients of interest with respect to the experiment are those associated with the variables $Treat_d$, $DirectEmail_d$, and $Treat_d \times DirectEmail_d$. In particular, the coefficient γ_1 corresponds to the average change in the response rate from call-to-serve emails with the peer comparison treatment. The coefficient γ_2 corresponds to the average change in the response rate from additional call-to-serve emails sent directly to faculty. Finally, the coefficient γ_3 corresponds to the average change in the response rate from peer comparison emails sent directly to faculty. The matrix X is a set of control variables, which include indicator variables corresponding to the size quartile of department d ($SizeQ_d$), indicator variables corresponding to the service quartile of department d among its peers of similar size ($ServeQ_d$), and all the interactions of $SizeQ_d$ and $ServeQ_d$. We estimate equation 1 using both a linear probability model and a Probit model.

In addition to estimating average treatment effects, we fully interact equation 1 with the $ServeQ_d$ indicator variables in order to compute heterogeneity in treatment responses based on the service quartile revealed in the treatment email. We also investigate response heterogeneity by the gender of the faculty member and by whether the faculty member replied to the call-to-serve email in the previous year. We do this by estimating our preferred specification on the corresponding sub-samples of faculty members.

4 Results

4.1 Average Peer Comparison Treatment Effects and Heterogeneity by Service Quartile

While the above summary statistics are illuminating, we pursue a more formal analysis of the call-to-serve response rate changes caused by the different treatments. Table 5 reports the average treatment effects (columns 1–4), as well as the heterogeneous treatment effects by service quartile groups (columns 5–6), using specifications of equation 1. The odd numbered columns report the estimates from a linear regression, while the even numbered columns report the estimates from a Probit regression. The dependent variable in all columns is equal to one if the faculty member responded to the call-to-serve in 2014-15 and equal to zero otherwise. We include a set of indicator variables for the size quartile of the faculty member's department—*Size Q1* is the omitted reference size quartile—and a set of indicators for the service quartile of the faculty member's department—*Serve Q1* is the omitted reference service quartile—as well as their interactions. Faculty members in the highest serving pre-treatment departments (*Serve Q4*) are omitted from this analysis. Standard errors in parentheses are clustered at the department level.

Starting with the average effects in columns 1–4, while the *Treat* and *Treat* × *Direct* coefficients are all positive—which means on average the treatment emails increase the probability of faculty responding relative to the control emails—these estimates are not statistically different from zero. Furthermore, receiving a direct email, regardless if it is a control or treatment email, does not increase the probability of responding on average, as shown by the positive yet insignificant coefficients for *Direct Email* in columns 3 and 4. Thus, we cannot conclude direct emails are more or less effective than emails to the chair. This null result could reflect the signal from chair emails being stronger in some cases and weaker in other cases, depending on the (unobserved) way in which chairs choose to forward emails. If some chair emails are effective while others are not, they may cancel each other.

Moving now to the heterogeneous treatment effects in columns 5 and 6, breaking up the effects by the service quartile yields more nuanced results. The treatment effect for the lowest service quartile (*Serve Q1*) is positive and statistically significant, which can be seen by the coefficients on *Treat* in columns 5–6 (since *Serve Q1* is the omitted service quartile indicator). This means that the probability that faculty belonging to departments in the lowest service quartile respond to the call-to-serve increases when their chair receives a treatment email disclosing their low service rank. Conversely, the treatment email has a slightly negative effect for departments in the middle service quartiles, as shown by the coefficients on *Treat* $\times$ *Serve Q2* and *Treat* $\times$ *Serve Q3*, which are negative, significant, and larger in magnitude than the coefficients on *Treat*.

In summary, we find that the effect of revealing quartile rank is non-linear, with faculty in the lowest service quartile increasing their response rate with treatment and faculty in the middle service quartiles decreasing their response rate with treatment. Also, we do not find that sending direct emails to faculty increases the response rate on average or heterogeneously by service quartile.

4.2 Gender Differences in Peer Comparison Treatment Effects

Next, we examine gender differences in the treatment effects using the four model specifications shown in Table 6. Column 1 replicates the OLS specification from column 5 of Table 5 with the addition of an indicator variable *Male*, equal to 1 if faculty member f is male. Columns 2 and 3 estimate the model separately for the 801 male faculty and the 325 female faculty. Lastly, column 4 fully interacts the model in column 1 with the indicator for *Male*, in order to test whether the gender differences found in column 2 and 3 are statistically significant.

We find several interesting results comparing men and women. First, simply controlling for gender reveals no differences in male and female response rates on average (column 1). Second, the treatment effect for faculty in *Serve Q1* is larger for women. In column 3, treated women in *Serve Q1* were 15.41 percentage points more likely to respond to the call-to-serve than control women in *Serve Q1.* While men in *Serve Q1* also have a positive coefficient on *Treat* (column 2), it is smaller in magnitude than the coefficient for women (0.1018 vs 0.1541) and it is not statistically different than zero, even though the sample size of men is more than double that of women. However, even though it appears the women are driving the treatment effect in *Serve Q1*, the fully interacted model in column 4 reveals that the 5 percentage point difference in treatment effects between genders is not statistically significant, due to the noisier estimate for men. Similarly, the treatment effect in *Serve Q2* is larger for women than men, but this difference is not statistically significant (column 4). On the other hand, column 4 shows that without treatment, men are more likely to respond than women to serve in *Serve Q1* (with a positive and significant coefficient on *Male*).

The third result of interest is that the treatment effect for faculty in *Serve Q3* is larger for men than women, with a negative and significant coefficient in the row labeled *Treat* $\times$ *Serve Q3* for men in column 2 and a positive and insignificant coefficient for women in column 3. In particular, treated men in *Serve Q3* were 20.45 percentage points less likely to respond to the call-to-serve than control men in *Serve Q3*. Moreover, column 4 reveals the coefficients on *Treat* $\times$ *Serve Q3* are statistically different between men and women.

In summary, we find the *Serve Q1* treatment effect is driven by women and the *Serve Q3* effect is driven by men. This finding is consistent with women being more motivated by below average peer information than men when responding to service calls, and men being more motivated by above average peer information. Vesterlund et al. (2015) show that in thinking about accepting the work-related requests, women more than men experience negative emotions (stress, anxiety, guilt), and worry about the negative repercussions of declining such request. Our results are congruous with these emotions being exacerbated if women learn they are behind their peers. However, our findings are also consistent with social learning, where women in low-serving departments make a Bayesian inference that they are under-serving relative to their peers, and men in higher serving departments make a Bayesian inference that they are over-serving and being free-ridden by their peers, since they could be spending their time in research. We also cannot rule out that women responded to the treatment in *Serve Q1* because treated chairs asked (independently of the call-to-serve email) female faculty to serve more than men.

An important contribution of our experimental design is that it allows us to look at gender differences both at baseline (i.e., pre-experiment) and due to the peer comparison treatment. In particular, our baseline response data allows us to relate our findings to previous research analyzing gender differences in standard calls for voluntary service, without peer comparison interventions. Babcock et al. (2017) examine response data from another large public university where each year the chair of the academic senate sends a call-to-serve email to all faculty members asking them to volunteer to serve on a faculty senate committee. This differs from our setting where the standard call-to-serve email is sent to department chairs and requests the chairs forward the email to their departments.

Babcock et al. (2017) find that female faculty are significantly more likely than male faculty to respond to call-to-serve emails saying they would volunteer to be on a committee. In Table 7, we replicate the Probit model of Babcock et al. (2017) using the 2013-14 baseline response data (column 1), 2014-15 response data for the control departments (column 2), and 2014-15 response data for departments receiving the direct email (column 3). Across all three samples, we do not find that female faculty respond significantly more than male faculty to the call-to-serve emails. If anything, the coefficients for female faculty are all negative.

Additionally, the two universities differ in the representation of women in the entire faculty and on the academic senate committees. In the year of their call-to-serve data, Babcock et al. (2017) find that women constituted 24.7% of faculty and 37.5% of committee positions. In our baseline year, women comprised 30.1% of faculty and 32.7% of committee positions. Thus in the Babcock et al. (2017) sample, women are substantially over-represented on the committees, while in our sample the gender composition of committees is roughly proportional to the whole faculty.

Unlike the gender differences in treatment effects we find in Table 6, we find no gender differences in responding at baseline. Furthermore, as also shown in summary Table 3, we cannot conclude that women are more likely to serve on committees.[15] The fact that our baseline results counter those of Babcock et al. (2017) indicates that gender differences may be sensitive to observed factors, such as how the call-to-serve email is sent, and unobserved factors, such as institutional norms. Institution norms may include faculty beliefs on how important senate service is for promotion past tenure. Some studies have shown that gender gaps get smaller when the rewards to a task become larger (Mulligan and Rubinstein, 2008; Petrie and Segal, 2015). The university in our setting emphasizes the importance of service for promotion in their call-to-serve emails; this may not have been the case in the Babcock et al. (2017) setting.

The differences in our results may also be reconciled by Croson and Gneezy (2009) and Jones and Linardi (2014), who find that women are neither more nor less socially oriented, but that social preferences of women are more situationally specific and malleable than those of men. Babcock et al. (2017) argue that the mechanism driving women more than men to receive requests to volunteer and accept requests to volunteer is the expectation that they are more likely than men to volunteer when *someone needs to do so* (i.e., women are more likely to step up to the plate). Since our field experiment provides information about a department's relative service and its need for volunteers, the result that women in the lowest serving departments become more likely to volunteer is consistent with women in these departments stepping up to the plate more than men when someone needs to take one for the team (or with chairs asking women more than men to step up to the plate when

[15] Another difference between Babcock et al. (2017) and this study is the average response rate to the call-to-serve email. The average pre-experiment response rate at the public university in this study is 8.5%, while Babcock et al. (2017) find a volunteer response rate of 3.7% at the university in their study.

there is a need).

In summary, even though we find that social comparisons are effective at increasing service responses among faculty in the lowest serving departments, these effects are driven by female faculty responses, be it from the treatment response of individual faculty or chairs. This was particularly interesting since women and men at this university did not differ in their response rates to taking on service before the experiment. Our findings suggest that the information faculty and chairs have about their peers' service and how faculty are asked to volunteer could explain some of the gender differences in performing low-promotability tasks, and consequently, could be an important component in the career trajectory of women relative to men.

4.3 Heterogeneity in Peer Comparison Treatment Effects by Previous Responses

In Table 8, we examine if the treatment effects vary by whether the faculty member replied to the call-to-serve in the previous year. We replicate the OLS specification from column 5 of Table 5 separately for the 83 faculty that replied to the 2013-14 call-to-serve (shown in column 1) and for the 1043 faculty that did not reply (shown in column 2). Faculty in *Serve Q4* are once again excluded from the analysis. In column 1, faculty that replied to the call-to-serve in the previous year are more likely to respond again if they receive a treatment email revealing their department is in the lowest service quartile and are less likely to respond again if the email reveals their department is in the above median service quartile. Faculty that replied to the call-to-serve in the previous year are also much more likely to respond again if they receive a direct email, regardless if it is a treatment or control email, as shown by the large and significant coefficient for *Direct Email*.

In column 2, faculty that did not reply to the call-to-serve in the previous year are also more likely to respond if they receive an email saying their department is in the lowest service quartile and less likely to respond if revealed to be above median, however, the coefficients are smaller in magnitude than those in column 1 and are no longer statistically different than zero. Moreover, unlike in column 1, receiving a direct email does not impact the response of faculty that did not respond in the previous year. Thus while this experiment was successful in redistributing serving from high to low service departments, these changes were stronger for the faculty members who had already engaged in service. If the institution wants to increase service among those who have not volunteered before, these interventions may not be desirable.

5 Conclusion

This paper provides a first step towards understanding if peer comparisons affect the behavior of those called to serve on faculty senate committees. We implement a randomized field experiment at a large public university to examine whether revealing a department's service ranking among its peers, in terms of previous service on faculty committees, affects future response rates of faculty members to calls to serve. We find that revealing a low service ranking among peers leads to significantly more service responses than disclosing a near-median service ranking. Yet beyond informing department heads of their departments' service rank, directly informing individual faculty members does not have an additional effect on service responses.

In terms of heterogeneous effects by gender, we extend the findings in Babcock et al.

(2017). While contrary to Babcock et al. (2017), we do not find gender differences in the response rate to the standard call-to-serve emails pre-experiment, we estimate gender differences in the call-to-serve response rate when there is a peer comparison treatment. In particular, we find that women in departments revealed to be in the lowest service quartile have a larger percentage point increase in their response rate than their male colleagues, while men in departments revealed to be above median have a larger percentage point decrease in their response rate than their female colleagues. While we cannot rule out that these treatment effects are driven by the behavior of chairs, rather than individual faculty, these results imply that information about peer behavior and *how* women and men are asked to do low-promotability tasks may have large impacts on career trajectories.

This is the first paper to examine whether responses to calls to serve in university governance is influenced by internal service rankings. To make general conclusions about the determinants of voluntary service, systematic data collection and analysis are required on volunteer service rates and call-to-serve response rates at other institutions. Future work could also consider making the department specific disclosed information not only available to each particular unit, as in this paper's design, but also available to all units. This could be done in the form of a disclosed service performance "Top Service" list. Additional investigations could test whether the results are heterogeneous along other dimensions, such as age or how long someone has worked at the institution. Finally, while the context of this paper is higher education, the insights can be taken to other settings that suffer from volunteer fatigue. For instance, future work could explore whether peer comparisons redistribute service loads in non-profit sports clubs, religious organizations, and Parent-Teacher Associations that rely on member-volunteers to operate. One final step left unanswered is to disentangle the mechanisms behind our results. We find that revealing peer ranking increases response rates but we cannot say exactly why. Future work should identify whether the mechanism behind the different treatment effects is due to differences in pro-social behavior or whether it is due to differences in the ability to process and say no to calls to serve in low-promotion tasks.

References

Allcott, H. (2011). Social Norms and Energy Conservation. *Journal of Public Economics 95*(9), 1082–1095.

Ayres, I., S. Raseman, and A. Shih (2013). Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. *Journal of Law, Economics, and Organization 29*(5), 992–1022.

Babcock, L., M. Recalde, and L. Vesterlund (2017). Gender Differences in the Allocation of Low-Promotability Tasks: The Role of Backlash. *American Economic Review: Papers & Proceedings 107*(5), 131–35.

Babcock, L., M. P. Recalde, L. Vesterlund, and L. Weingart (2017). Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability. *American Economic Review 107*(3), 714–747.

Blackman, A., S. Afsah, and D. Ratunanda (2004). How Do Public Disclosure Pollution Control Programs Work? Evidence from Indonesia. *Human Ecology Review 11*(3), 235–246.

Card, D., A. Mas, E. Moretti, and E. Saez (2012). Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. *American Economic Review 102*(6), 2981–3003.

Cialdini, R. and N. Goldstein (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology 55*, 591–621.

Croson, R. and U. Gneezy (2009). Gender Differences in Preferences. *Journal of Economic Literature 47*(2), 448–474.

Frey, B. S. and S. Meier (2004). Social Comparisons and Pro-Social Behavior: Testing "Conditional Cooperation" in a Field Experiment. *American Economic Review 94*(5), 1717–1722.

Gerber, A. S. and T. Rogers (2009). Descriptive Social Norms and Motivation to Vote: Everybody's Voting and So Should You. *Journal of Politics 71*(1), 178–191.

Goldstein, N. J., R. B. Cialdini, and V. Griskevicius (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research 35*(3), 474–482.

Jin, G. and P. Leslie (2003). The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *Quarterly Journal of Economics 118*(2), 409–451.

Jones, D. and S. Linardi (2014). Wallflowers: Experimental Evidence of an Aversion to Standing Out. *Management Science 60*(7), 1757–1771.

Misra, J., J. H. Lundquist, and A. Templer (2012). Gender, Work Time, and Care Responsibilities Among Faculty. *Sociological Forum 27*(2), 300–323.

Mitchell, S. M. and V. L. Hesli (2013). Women Don't Ask? Women Don't Say No? Bargaining and Service in the Political Science Profession. *PS: Political Science & Politics 46*(2), 355–369.

Mulligan, C. B. and Y. Rubinstein (2008). Selection, investment, and women's relative wages over time. *Quarterly Journal of Economics 123*(3), 1061–1110.

Penner, L. A., J. F. Dovidio, J. A. Piliavin, and D. A. Schroeder (2005). Prosocial Behavior: Multilevel Perspectives. *Annual Review of Psychology 56*, 365–392.

Petrie, R. and C. Segal (2015). Gender differences in competitiveness: The role of prizes. *GMU Working Paper in Economics 14*(47), 1–33. Available at SSRN: https://ssrn.com/abstract=2520052 or http://dx.doi.org/10.2139/ssrn.2520052.

Porter, S. R. (2007). A Closer Look at Faculty Service: What Affects Participation on Committees? *The Journal of Higher Education 78*(5), 523–541.

Smith, S., F. Windmeijer, and E. Wright (2015). Peer Effects in Charitable Giving: Evidence from the (Running) Field. *The Economic Journal 125*(585), 1053–1071.

Vesterlund, L., L. Babcock, and L. Weingart (2015). Breaking the Glass Ceiling with "No": Gender Differences in Declining Requests for Non-Promotable Tasks. Working paper, Online.

Table 1: **Service Participation Summary Statistics by Department Size Quartile**

	Size Q1	Size Q2	Size Q3	Size Q4
Number of Faculty in Department	up to 9.5	up to 16	up to 27.5	up to 80.5
Number of Service Eligible Faculty in Department	[2.5,7.5]	(7.5,12]	(12,24.5]	(24.5,72.5]
Number of Departments	13	19	15	17
Average Service Eligible Participation in 2013-14	26.03%	32.51%	27.14%	22.85%
Median Service Eligible Participation in 2013-14	16.67%	31.58%	24.32%	24.24%
Maximum Service Eligible Participation in 2013-14	75.00%	75.00%	57.73%	33.71%

Size Q1 is comprised of the smallest departments—25^{th} percentile or lower with respect to the number of service eligible faculty—and *Size Q4* is comprised of the largest departments—75^{th} percentile or higher with respect to the number of service eligible faculty. The service eligible pool consists of faculty that are Associate Professors or higher. Source: Faculty 2013-14 census roster by department.

Table 2: **Number of Departments in the Treatment and Control Groups, by Service Participation Quartile and Size Quartile**

	Serve Q1		Serve Q2		Serve Q3		Serve Q4
	control	treat	control	treat	control	treat	no treat
Size Q1	1	4	0	0	4	0	4
Size Q2	1	3	3	3	2	2	5
Size Q3	2	2	1	2	3	2	3
Size Q4	3	2	2	2	2	2	4

Serve Q1 is the lowest 25^{th} percentile of departments and *Serve Q4* is the 75^{th} or higher percentile of departments with respect to pre-treatment service participation rates (among departments of the same size). *Size Q1* is comprised of the smallest departments (25^{th} percentile or lower) and *Size Q4* is comprised of the largest departments (75^{th} percentile or higher), with respect to the number of service eligible faculty.

Table 3: **Sample Characteristics, For All Departments and by Service Quartile and Treatment Group**

	(1) All	(2) Serve Q1	(3) Serve Q2	(4) Serve Q3	(5) Serve Q4	(6) Chair Email control	(7) Chair Email treat	(8) Faculty Email control	(9) Faculty Email treat
Number of:									
Departments	64	18	13	17	16	18	18	6	6
Faculty	1501	420	385	321	375	455	470	117	84
2013-14 Department Averages:									
Number of Faculty	23.45	23.33	29.62	18.88	23.44	25.28	26.11	19.50	14.00
	(18.60)	(19.71)	(21.51)	(12.42)	(20.53)	(21.17)	(19.09)	(12.18)	(4.47)
Female Share of Faculty	0.35	0.31	0.33	0.36	0.39	0.35	0.32	0.31	0.36
	(0.19)	(0.18)	(0.15)	(0.17)	(0.24)	(0.19)	(0.17)	(0.11)	(0.18)
Share Replied to Serve	0.09	**0.03**	0.08	0.12	0.13	0.11	0.05	0.09	0.04
	(0.09)	(0.04)	(0.05)	(0.12)	(0.10)	(0.12)	(0.04)	(0.05)	(0.05)
Female Share Replied to Serve	0.07	0.04	0.05	0.10	0.09	0.09	0.05	0.08	0.04
	(0.11)	(0.13)	(0.09)	(0.11)	(0.12)	(0.13)	(0.09)	(0.14)	(0.10)
Male Share Replied to Serve	0.10	**0.03**	0.10	0.14	0.14	0.13	0.05	0.09	0.05
	(0.13)	(0.05)	(0.07)	(0.19)	(0.14)	(0.19)	(0.06)	(0.07)	(0.06)
Share Serving	0.18	**0.07**	**0.19**	**0.20**	0.27	0.16	0.16	**0.19**	0.05
	(0.11)	(0.06)	(0.08)	(0.09)	(0.09)	(0.12)	(0.09)	(0.05)	(0.04)
Share Women Serving	0.19	**0.08**	0.19	**0.15**	0.35	0.12	0.16	0.22	0.04
	(0.20)	(0.12)	(0.17)	(0.13)	(0.26)	(0.11)	(0.15)	(0.20)	(0.10)
Share Men Serving	0.17	**0.06**	0.21	0.22	0.21	0.20	0.15	0.18	0.06
	(0.13)	(0.07)	(0.11)	(0.16)	(0.11)	(0.19)	(0.09)	(0.13)	(0.05)
Trend in Service Participation:									
2005–2014	0.00	**-0.02**	0.00	0.00	0.01	-0.01	-0.01	0.00	-0.02
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Trend in Women Percentage:									
2008–2014	0.00	0.01	-0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)

Standard deviations are in parentheses, except in the bottom row which presents the standard errors for the trend in service participation. *Serve Q1* is the lowest quartile of departments with respect to pre-treatment service participation (among departments of the same size) and *Serve Q4* is the highest quartile of departments. Bold text in columns 2 to 4 indicates the average for the *Serve Q1*, *Q2*, and *Q3* departments is statistically different (with a p-value ≤ 0.10) from the average for the *Serve Q4* departments. Bold text in column 6 and 8 indicates the average for the treated departments is statistically different (with a p-value ≤ 0.05) from the average for the corresponding control departments in columns 7 and 9. Italics text in all columns indicates the average for male faculty is statistically different (p-value ≤ 0.05) from the average for female faculty.

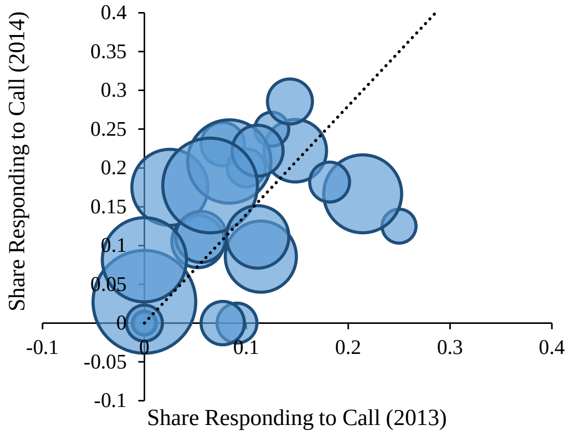
Table 4: **Call-to-Serve Responses for 2014-15**

	Replied for 2014-15		Did not reply for 2014-15	
Chair Only Email				
Treated (N=470)	74	16%	396	84%
Control (N=455)	58	13%	397	87%
Direct Faculty Email				
Treated (N=84)	19	23%	65	77%
Control (N=117)	22	19%	95	81%
Serve Q1				
Treated (N=198)	37	19%	161	81%
Control (N=222)	20	9%	202	91%
Serve Q2				
Treated (N=227)	35	16%	183	84%
Control (N=167)	30	18%	137	82%
Serve Q3				
Treated (N=138)	21	15%	117	85%
Control (N=183)	30	16%	153	84%
Male Faculty				
Treated (N=407)	72	18%	335	82%
Control (N=394)	57	14%	337	86%
Female Faculty				
Treated (N=147)	21	14%	126	86%
Control (N=178)	23	13%	155	87%
Did Not Respond to Call in 2013-14				
Treated (N=520)	75	14%	445	86%
Control (N=523)	55	11%	468	89%
Responded to Call in 2013-14				
Treated (N=34)	18	53%	16	47%
Control (N=49)	25	51%	24	49%

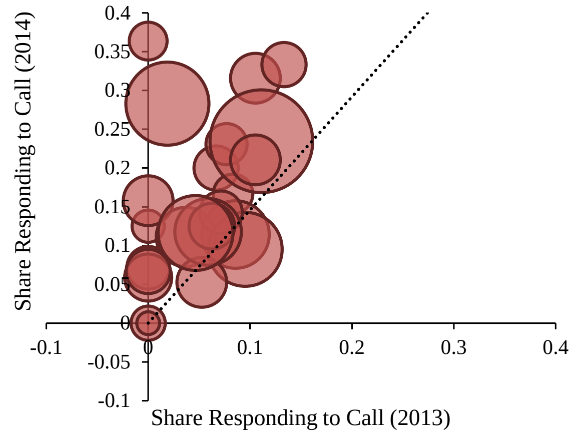
Source: 2014-15 call-to-serve reply database. Note: This table reports the number and share of 2014-15 call-to-serve responses by treatment and control group and by pre-treatment service quartile. *Serve Q1* is the lowest quartile of departments with respect to pre-treatment service participation (among departments of the same size) and *Serve Q4* is the highest quartile of departments. The 16 departments in the highest serving quartile (*Serve Q4*) are excluded from the control group in the top half of the table.

Figure 1: **Share of Treatment and Control Department Faculty Responding to the Call-to-Serve for 2013-14 (pre-treatment) and for 2014-15 (post-treament)**

(a) Control Departments



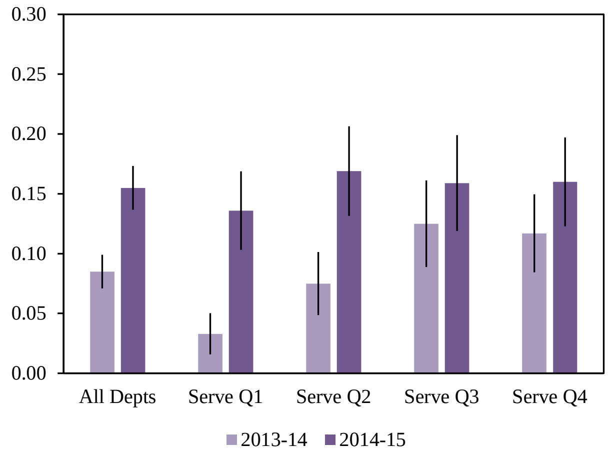
(b) Treated Departments



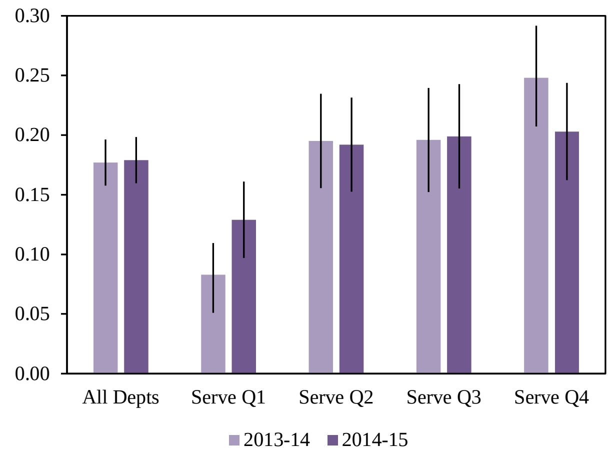
Source: 2013-14 and 2014-15 call-to-serve reply database. Note: Control departments depicted in blue. Treated departments depicted in red. Circle size corresponds to the size of each department.

Figure 2: **Share of Faculty Responding to the Call-to-Serve Email and Serving on a Committee, for All Departments and by Pre-Treatment Service Quartile**

(a) Share Responding to Call-to-Serve Email



(b) Share Serving on Academic Senate Committee



Source: 2013-14 and 2014-15 call-to-serve service database. Note: The pre-treatment year is 2013-14 and the treatment year is 2014-15. Vertical lines represent 95% confidence intervals.

Table 5: **Average Treatment Effects and Heterogeneity by Service Quartile**

	(1) OLS	(2) Probit	(3) OLS	(4) Probit	(5) OLS	(6) Probit
Treat	0.0351	0.1515	0.0310	0.1376	0.1138*	0.5328*
	(0.0278)	(0.1170)	(0.0319)	(0.1400)	(0.0568)	(0.2741)
Direct Email			0.0385	0.1686	0.0104	0.1259
			(0.0290)	(0.1231)	(0.0560)	(0.2760)
Treat × Direct			0.0519	0.1979	0.0527	0.1980
			(0.0501)	(0.2042)	(0.0891)	(0.3994)
Treat × Serve Q2					-0.1301*	-0.6002**
					(0.0649)	(0.3022)
Treat × Serve Q3					-0.1434**	-0.6664**
					(0.0701)	(0.3233)
Direct × Serve Q2					-0.0185	-0.1601
					(0.0605)	(0.2892)
Direct × Serve Q3					0.0874	0.2636
					(0.0705)	(0.3456)
Treat × Direct × Serve Q2					-0.0601	-0.2204
					(0.0991)	(0.4338)
Treat × Direct × Serve Q3					0.0620	0.1863
					(0.0970)	(0.4312)
Constant	0.1077	-1.2234***	0.0636	-1.4448***	0.0045	-1.7888***
	(0.0831)	(0.3751)	(0.0642)	(0.3263)	(0.0752)	(0.4010)
N	1126	1126	1126	1126	1126	1126
Service × Size Interactions	X	X	X	X	X	X
R-sq	0.0119		0.0155		0.0246	
LL	-442.4704	-476.1165	-440.4499	-474.1369	-435.2249	-468.7185

Dependent variable in all columns is equal to one if the faculty member responded to the call-to-serve for 2014-15 and equal to zero otherwise. Standard errors in parentheses are clustered at the department level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Odd-numbered columns have the estimates from a linear regression, while even numbered columns have the estimates from a Probit regression. *Serve Q1* is the lowest quartile of departments with respect to pre-treatment service participation (among departments of the same size) and *Serve Q4* is the highest quartile of departments. Faculty in untreated *Serve Q4* departments are not included in this analysis. *Serve Q1* is the reference (omitted) category in all columns. Though not reported in the table, regressions in columns 5 and 6 include all size and service participation quartile dummies and their lower order interactions.

Table 6: **Heterogeneity in Treatment Effects by Gender (*OLS*)**

	(1) Whole Sample	(2) Male Sample	(3) Female Sample	(4) Whole Sample, Fully Interacted by Gender
Treat	0.1129**	0.1018	0.1541***	0.1541***
	(0.0539)	(0.0752)	(0.0410)	(0.0403)
Direct Email	0.0126	-0.0048	0.1096	0.1096
	(0.0553)	(0.0685)	(0.0727)	(0.0716)
Treat × Direct	0.0552	0.1009	-0.0624	-0.0624
	(0.0884)	(0.1143)	(0.0770)	(0.0759)
Treat × Serve Q2	-0.1334**	-0.1071	-0.2313***	-0.2313***
	(0.0617)	(0.0862)	(0.0576)	(0.0568)
Treat × Serve Q3	-0.1438**	-0.2045**	0.0665	0.0665
	(0.0674)	(0.0904)	(0.0768)	(0.0757)
Direct × Serve Q2	-0.0247	-0.0641	0.0418	0.0418
	(0.0599)	(0.0808)	(0.1179)	(0.1162)
Direct × Serve Q3	0.0863	0.1592*	-0.0655	-0.0655
	(0.0694)	(0.0836)	(0.1067)	(0.1051)
Treat × Direct × Serve Q2	-0.0645	-0.0653	-0.1662	-0.1662
	(0.0977)	(0.1341)	(0.1297)	(0.1278)
Treat × Direct × Serve Q3	0.0518	0.0777	-0.1077	-0.1077
	(0.0958)	(0.1262)	(0.1045)	(0.1029)
Male	0.0327			0.2078***
	(0.0270)			(0.0760)
Treat × Male				-0.0524
				(0.0709)
Direct Email × Male				-0.1144
				(0.0902)
Treat × Direct × Male				0.1633*
				(0.0880)
Treat × Serve Q2 × Male				0.1242
				(0.0959)
Treat × Serve Q3 × Male				-0.2711***
				(0.0977)
Direct × Serve Q2 × Male				-0.1059
				(0.1536)
Direct × Serve Q3 × Male				0.2248*
				(0.1245)
Treat × Direct × Serve Q2 × Male				0.1009
				(0.1785)
Treat × Direct × Serve Q3 × Male				0.1853
				(0.1248)
Constant	-0.0145	0.0846	-0.1231**	-0.1231**
	(0.0725)	(0.0948)	(0.0580)	(0.0573)
N	1126	801	325	1126
Service × Size Interactions	X	X	X	
Service × Size × Male Interactions				X
R-sq	0.0246	0.0405	0.0824	0.0524

Dependent variable in all columns is equal to one if the faculty member responded to the call-to-serve for 2014-15 and equal to zero otherwise. Standard errors in parentheses are clustered at the department level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. *Serve Q1* is the lowest quartile of departments with respect to pre-treatment service participation (among departments of the same size) and *Serve Q4* is the highest quartile of departments. Faculty in untreated *Serve Q4* departments are not included in this analysis. *Serve Q1* is the reference (omitted) category in both columns. Though not reported in the table, regressions include all size and service participation quartile dummies and their lower order interactions.

Table 7: **Probability of Responding to the Call-to-Serve without Treatment (*Probit*)**

	(1)	(2)	(3)
	2013-14	2014-15	2014-15
		Untreated	Direct Email
	All Depts	Depts	Depts
Female	-0.0164	-0.0291	-0.0529
	(0.0149)	(0.0254)	(0.0583)
Associate Professor	0.1436***	0.0708	0.3158**
	(0.0501)	(0.0517)	(0.1443)
Full Professor	0.0848***	0.0906***	0.1562*
	(0.0223)	(0.0345)	(0.0861)
STEM	-0.0180	-0.0935***	0.0259
	(0.0144)	(0.0246)	(0.0573)
N	1501	830	117
ll	-425.8559	-329.4450	-97.9967

This table replicates Table 1 in Babcock et al. (2017) using data from our study's faculty census matched to data from our 2013-14 and 2014-15 call-to-serve reply database. This table presents marginal effects from a Probit regression. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dependent variable is equal to one if the faculty member replied to the call-to-serve and equal to zero otherwise. Column 1 includes response data for 2013-14 (pre-treatment) from all departments, column 2 includes data from the un-treated department in 2014-15, and column 3 includes data from direct email departments in 2014-15.

Table 8: **Heterogeneity in Treatment Effects by Previous Responses to the Call-to-Serve (*OLS*)**

	(1) Served in 2013-14	(2) Did Not Serve in 2013-14
Treat	0.3125***	0.0984
	(0.1002)	(0.0614)
Direct Email	0.8750***	-0.0425
	(0.1009)	(0.0522)
Treat × Direct	-0.1875	0.0975
	(0.1879)	(0.0830)
Treat × Serve Q2	-0.1724	-0.1236*
	(0.1687)	(0.0664)
Treat × Serve Q3	-0.4635**	-0.0685
	(0.1841)	(0.0692)
Direct × Serve Q2	-0.3960*	0.0324
	(0.2295)	(0.0550)
Direct × Serve Q3	0.0998	0.0160
	(0.1072)	(0.0725)
Treat × Direct × Serve Q2	-0.0374	-0.1202
	(0.2425)	(0.0876)
Treat × Direct × Serve Q3	-0.1678	0.0696
	(0.2397)	(0.0882)
Constant	0.0126	0.0224
	(0.0612)	(0.0764)
N	83	1043
Service × Size Interactions	X	X
R-sq	0.3407	0.0227

Dependent variable in all columns is equal to one if the faculty member responded to the call-to-serve for 2014-15 and equal to zero otherwise. Standard errors in parentheses are clustered at the department level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. *Serve Q1* is the lowest quartile of departments with respect to pre-treatment service participation (among departments of the same size) and *Serve Q4* is the highest quartile of departments. Faculty in untreated *Serve Q4* departments are not included in this analysis. *Serve Q1* is the reference (omitted) category in both columns. Though not reported in the table, regressions include all size and service participation quartile dummies and their lower order interactions.

Figure Appendix.1: **Emails sent to Chairs and to All Senate Members**

===BEGIN CHAIR EMAIL:
Dear Chair (Dean) XXX,
We would like to thank your department for its ongoing participation in the activities of the [University's] Academic Senate. [The committee that recruits service] will soon begin to recommend faculty for service on the [Academic Senate]. [The University's] principle of shared governance contributes importantly to the excellence of this [university]. Shared governance depends, however, on the willingness of faculty to exercise it. There are two other, more concrete reasons why it is in the interest of your faculty to undertake Academic Senate service. First, such service gives your department a place at the table and assures that its interests are represented in decisions affecting our lives as teachers, researchers, and employees. Second, as the [Academic Senate and Administration] have forcefully emphasized, participation in Academic Senate and administrative committees is an important consideration in advancement and promotion, especially after tenure is received.
==Only for Treatment Departments==After reviewing the service participation data for 64 campus units, we have noted that your department is in the bottom 25% when compared to other units of similar size.
Please forward this message to your faculty to encourage them to submit their committee preferences directly. They can sign up by following this link:<<LINK HERE>>
Once again, we would be very grateful if you would encourage your colleagues to respond. Thank you again for your continued support of the [University's] Academic Senate!
===END OF EMAIL

===BEGIN DIRECT FACULTY EMAIL:
Dear Colleague,
This is a reminder for you to respond to the call to serve. [The committee that recruits service] will soon begin to recommend faculty for service on the [Academic Senate], and I invite you to serve. [The University's] principle of shared governance contributes importantly to the excellence of this campus and university. Shared governance depends, however, on the willingness of faculty to exercise it.
There are two other, more concrete reasons why it is in your interest to undertake Academic Senate service. First, such service gives your department a place at the table and assures that its interests are represented in decisions affecting our lives as teachers, researchers, and employees. Second, as the [Academic Senate and Administration] have forcefully emphasized, participation in Academic Senate and administrative committees is an important consideration in advancement and promotion, especially after tenure is received.
==Only for Treatment Departments==After reviewing the service participation data for 64 campus units, we have noted that your department is in the bottom 25% when compared to other units of similar size.
We would therefore like to encourage you to submit your committee preferences directly. You can sign up by following this link: <<<LINK HERE>>>
Thank you again for your continued support of the [University's] Academic Senate!
===END OF EMAIL

Note: Edited Chair Email and Faculty Email are provided here. Treated departments receive the entire email content. Control departments do not receive text marked in blue.